U.S. DEPARTMENT OF COMMERCE
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
NATIONAL WEATHER SERVICE
SYSTEMS DEVELOPMENT OFFICE
TECHNIQUES DEVELOPMENT LABORATORY

TDL Office Note 75-10

CURRENT STATUS OF PROBABILITY OF PRECIPITATION AMOUNT (POPA) FORECASTING

Robert J. Bermowitz and Edward A. Zurndorfer

December 1975

Current Status of Probability of Precipitation Amount (POPA) Forecasting

by

Robert J. Bermowitz and Edward A. Zurndorfer

INTRODUCTION

Probability of precipitation amount (POPA) and categorical forecasts of precipitation amount are being supplied as guidance to the Quantitative Precipitation Branch of the National Meteorological Center (NMC). Forecasts for the categories $\geq$ .25, $\geq$ .50, $\geq$ 1.0, and $\geq$ 2.0 inches are made with use of the Model Output Statistics technique (Glahn and Lowry, 1972) for the following projections: 12-36, 36-60, 12-24, 24-36, 36-48, 12-18, 18-24, and 24-30 hr after 0000 GMT, and 24-48, 12-18, 18-24, and 24-30 hr after 1200 GMT. Forecasts for the 12-hr periods are not supplied to NMC; they will be used as part of the computer worded forecast. Also, forecasts for the category $\geq$ 2.0 inch are not made for the 6-hr periods because of the extreme rarity of this event. Our purpose in writing this note is to discuss the current status of this guidance and present some comparative verification of the POPA forecasts.

There are currently two POPA systems in operation. The first, called POPA2, was put into operation during February 1975 and uses as predictors forecast fields from the primitive equation (PE) (Shuman and Hovermale, 1968) and trajectory (Reap, 1972) models. Forecasts in both probabilistic and categorical form are available for all projections listed above (also see Fig. 1). The second system, called POPA1, was put into operation during the cold season (October - March) of 1975-1976. It is an early guidance forecast package available about 2 hours earlier than POPA2 and uses forecast fields from only the limited area fine mesh (LFM) model (Howcroft and Desmaris, 1971) as predictors. Probabilistic and categorical forecasts are available for the 12-18 and 18-24 hr projections after both 0000 and 1200 GMT (Fig. 1).

POPA 1974-1975

The 1974-1975 POPA2 system consisted of generalized operator equations for the 6 regions shown in Fig. 2. These regions were derived by a subjective analysis of the frequency of occurrence of observed precipitation amounts for various forecast amounts from the PE model. For a given projection, equations for each category were developed with different "best" predictor sets (some predictors are, of course, repeated for different categories). Therefore, the final predictors screened for the category $\geq$ .25 inch were somewhat different than those screened for the category $\geq$ 2.0 inches. While this may seem intuitively more desirable than, say, screening the same predictor set for each category (this requires only one computer run), it is more expensive in terms of computer cost since separate computer runs must be made for each category. Generally speaking, the most important predictors were found to be precipitation amount, mean relative humidity from the surface to 500 mb, and precipitable water--all from the PE model.

Since categorical forecasts of precipitation amount are a part of the guidance product, we must transform the probability forecasts to categorical forecasts. A discussion of several methods of transforming POPA forecasts is given by Bermowitz (1975). The threat score[1] is the primary statistic used for verification of forecasts of precipitation amount at NMC. Therefore, we transformed the probability forecasts by maximizing the threat score. This requires calculation of a threshold probability for each category for each projection that will maximize the threat score for dichotomous forecasts of that category. The threshold probability for a category, say ≥ .25 inch, is a value that if exceeded by a probability forecast for that category, would result in a categorical forecast of ≥ .25 inch. If the threshold value is not exceeded, the categorical forecast would be < .25 inch.

We maximized the threat score by using one threshold value for the conterminous United States for each category. These values were computed by combining probability forecasts from all stations. Using this relatively large sample tends to produce stable threshold values. However, significant underforecasting of all categories ≥ .25 inch would occur over the dry regions if the probability forecasts are consistently below the threshold value.

VERIFICATION OF POPA2 1974-1975

To test the 1974-1975 system we performed a comparative verification of POPA2 categorical forecasts against those prepared subjectively at NMC. It is unlikely that the POPA guidance played a significant part in preparation of the subjective forecasts because of delays in running the computer program. Threat scores and biases[2] were computed at 232 cities for all categories for the 12-36 and 36-60 hr projections after 0000 GMT and for the 24-48 hr projection after 1200 GMT. In addition, threat scores and biases were computed at 230 cities for the category ≥ .25 inch for the 12-18 and 18-24 hr projections after 0000 and 1200 GMT. The category ≥ .25 inch is the only one verified since NMC did not record categorical forecasts greater than that for the 6-hr periods. We also verified LFM 6-hr forecasts of precipitation amount.

The period of verification for all forecasts was from February 19 - March 31, 1975. Because of missing forecasts, the sample size ranged from 32 to 39 days. For the 24-hr periods this consisted of about 1050 observations of ≥ .25 inch, 525 observations of ≥ .50 inch, 180 observations of ≥ 1.0 inch, and about 30 observations of ≥ 2.0 inches. For the 6-hr periods, there were about 250 observations of ≥ .25 inch.

The results of the verification of the forecasts for the 24-hr periods are shown in Table 1. The subjective forecasts had better threat scores for all categories for all projections except for the category ≥ .25 inch for the

---

[1] Threat score = H/(F+O-H) where H is the number of correct forecasts of a category and F and O are the number of forecasts and observations of that category.

[2] Bias is the number of forecasts of a category divided by the number of observations of that category. A categorical bias equal to 1 means unbiased forecasts of that category.

projections of 12-36 and 24-48 hr. However, in general, the subjective forecasts overforecast more than POPA2 except for the 24-48 hr projection where POPA had a higher bias.

Scores for the 6-hr periods are shown in Table 2. For the 12-18 hr projections, POPA2 and the subjective forecasts were about the same. The LFM had a better threat score than the others for this projection for the 1200 GMT cycle. For the 18-24 hr projection after 0000 GMT, the LFM again had the best threat score, followed by POPA2 and subjective in that order. However, the LFM overforecast the most. For the 18-24 hr projection after 1200 GMT, POPA2 had the best threat score followed by the LFM and subjective which were about the same. POPA2 and LFM overforecast the most for this projection.

POPA2 1975-1976

There are three differences between the POPA2 systems used during the cold seasons 1974-1975 and 1975-1976 other than the use of an additional cold season of dependent data (for a total of 5 seasons) for the 1975-1976 system. First of all, three regions were added for the 1975-1976 system for a total of nine, as shown in Fig. 3. We felt that some refinement in the distribution of regions would benefit the probability forecasts. Secondly, during development of the 1975-1976 equations, we screened the same predictor set for each category for a given projection rather than using different predictor sets. We did not feel that this method would seriously impair the accuracy of the probability forecasts (no testing was done); we were certain that use of it would result in reduced computer costs.

Third, the threat score was maxmized by region. A breakdown of the threat score and bias by region in the comparative verification indicated that the 1974-1975 system hardly ever made any categorical forecasts of $\geq$ .25 inch in region 5 (Fig. 2). This is an undesirable characteristic and is caused by using one threshold value for each category for the conterminous United States. Therefore, threshold values were determined for each region, a technique that alleviates the problem of serious underforecasting in the dry areas. Unfortunately, threshold values are derived on smaller data samples with this technique. This can make it difficult and, in some cases, impossible to arrive at an accurate value.

To try to obtain some indication of how the 1975-1976 system would perform on independent data, we made a comparative verification of POPA2 1975-1976 categorical forecasts against those prepared (1) by the POPA2 1974-1975 system, (2) subjectively at NMC, and (3) by the LFM for only the 6-hr forecast period. As in the previous verification, threat scores and biases were computed at 232 cities for all categories for the 12-36 and 36-60 hr projections after 0000 GMT and at 230 cities for the category $\geq$ .25 inch for the 12-18 hr projection after 0000 GMT. The period of verification was March 1975; the sample size ranged from 26 to 28 days.

Table 3 contains the results of this verification for the 24-hr forecast periods. They indicate improvement in threat score in nearly all categories for the POPA2 1975-1976 system when compared to the 1974-1975 system. The bias generally increased for the 1975-1976 system; however, it was designed to do so. If necessary, the bias could be lowered by using a higher threshold probability. However, some corresponding decrease in threat score would likely occur. Of importance is the fact that the 1975-1976 system is quite capable of making an acceptable number of correct forecasts of $\geq$ .25 inch over the dry areas of the Western United States. Indeed, forecasts of $\geq$ 1.0 inch were made there for 24-hr periods. Threat scores for the forecasts for 24-hr periods for the 1975-1976 system were still, with a few exceptions, below those of the subjective forecasts.

The results for the 12-18 hr forecasts after 0000 GMT are shown in Table 4. The threat score decreased slightly for the 1975-1976 system as compared to the 1974-1975 system; however, it was still slightly higher than that of the subjective forecasts. As was the case for 24-hr forecast periods, the bias also increased.

POPA1

The POPA1 system was developed to provide an early guidance package for the 12-18 and 18-24 hr projections for both 0000 and 1200 GMT cycles. Since predictors are used from only the LFM model, this package can be ready for use about 2 hr before the POPA2 guidance.

POPA1 equations and threshold probabilities to maximize the threat score were developed the same way as was the POPA2 system. Nine regions are also used for POPA1; they are slightly different than those used for POPA2. We determined these regions the same way as we did for POPA2, except we used LFM precipitation amount instead of PE precipitation amount.

We should point out that 3 cold seasons of dependent data were used to develop the POPA1 equations. A number of missing cases, especially in the first two seasons, reduced the sample size to a point that, for some categories, it was only about half that used for POPA2 developmental work. Therefore, the accuracy of the forecasts, primarily for the higher categories, could be questionable.

CONCLUDING REMARKS

A comparative verification on about one month of independent data indicated that the POPA2 1974-1975 forecasts were not as good as the subjective forecasts for the 24-hr forecast periods. However, the POPA forecasts did appear to have skill as measured by the threat score. As a matter of fact, POPA had nearly the same threat score as the subjective forecasts for (1) the category $\geq$ .25 inch for two of the 24-hr forecast periods and (2) for the category $\geq$ .50 inch for the 12-36 hr projection.

Skill of the POPA forecasts was more prominently indicated for the 6-hr forecast periods. They appeared somewhat better than the subjective forecasts. It is of interest, however, that the LFM forecasts were slightly better than either the POPA or subjective forecasts.

We feel that the 1975-1976 POPA2 system should provide improved guidance as indicated by the comparison with the 1974-1975 system. Certainly, many more forecasts of $\geq$ .25 inch will be made in the dry areas of the Western United States as a result of maximizing the threat score by region. Furthermore, we feel additional improvement, especially for the higher categories, will come about as we add additional years of developmental data. The larger sample size may facilitate evaluation of threshold probabilities that maximize the threat score in those regions where small samples now preclude an accurate evaluation. We also feel that improvement will result with the addition of climatological predictors. These could be in the form of frequencies of occurrence of various categories at stations.

ACKNOWLEDGMENTS

We would like to thank Mr. Julius Badner, former Chief, Quantitative Precipitation Branch, NMC, who supplied the data used in the verification. We appreciate the comments of the reviews: Dr. Harry R. Glahn and Mr. Roger A. Allen. Also, we thank the typist, Ms. Mary B. Battle.

REFERENCES

Bermowitz, Robert J., 1975: An application of model output statistics to forecasting quantitative precipitation. Monthly Weather Review, 103, 149-153.

Glahn, H. R., and D. A. Lowry, 1972: The use of model output statistics (MOS) in objective weather forecasting. Journal of Applied Meteorology, 11, 1203-1211.

Howcroft, J., and A. Desmaris, 1971: The limited area fine mesh (LFM) model. NWS Technical Procedures Bulletin, No. 67, 11 pp.

Reap, R. M., 1972: An operational three-dimensional trajectory model. Journal of Applied Meteorology, 11, 1193-1202.

Shuman, F. G., and J. B. Hovermale, 1968: An operational six-layer primitive equation model. Journal of Applied Meteorology, 7, 525-547.

```
PROJECTION (HOURS)                            00   06   12   18   24   30   36   42   48   54   60

VALID TIMES            00Z Cycle              00   06   12   18   00   06   12   18   00   06   12

                       12Z Cycle              12   18   00   06   12   18   00   06   12   18   00

                                               .    .    .    .    .    .    .    .    .    .    .

POPA1 GUIDANCE PACKAGE 6-HR PERIODS                      .- - -.- - -.

POPA2 GUIDANCE PACKAGE 6-HR PERIODS                      .- - -.- - -.- - -.

                      12-HR PERIODS                      .- - - - - -.- - - - - -.- - - - - -.

                      24-HR PERIODS (00Z)                .- - - - - - - - - - - -.- - - - - - - - - - - -.

                      24-HR PERIODS (12Z)                          .- - - - - - - - - - - -.
```

FIGURE 1. Forecast periods for POPA1 and POPA2 are indicated by dashed lines.

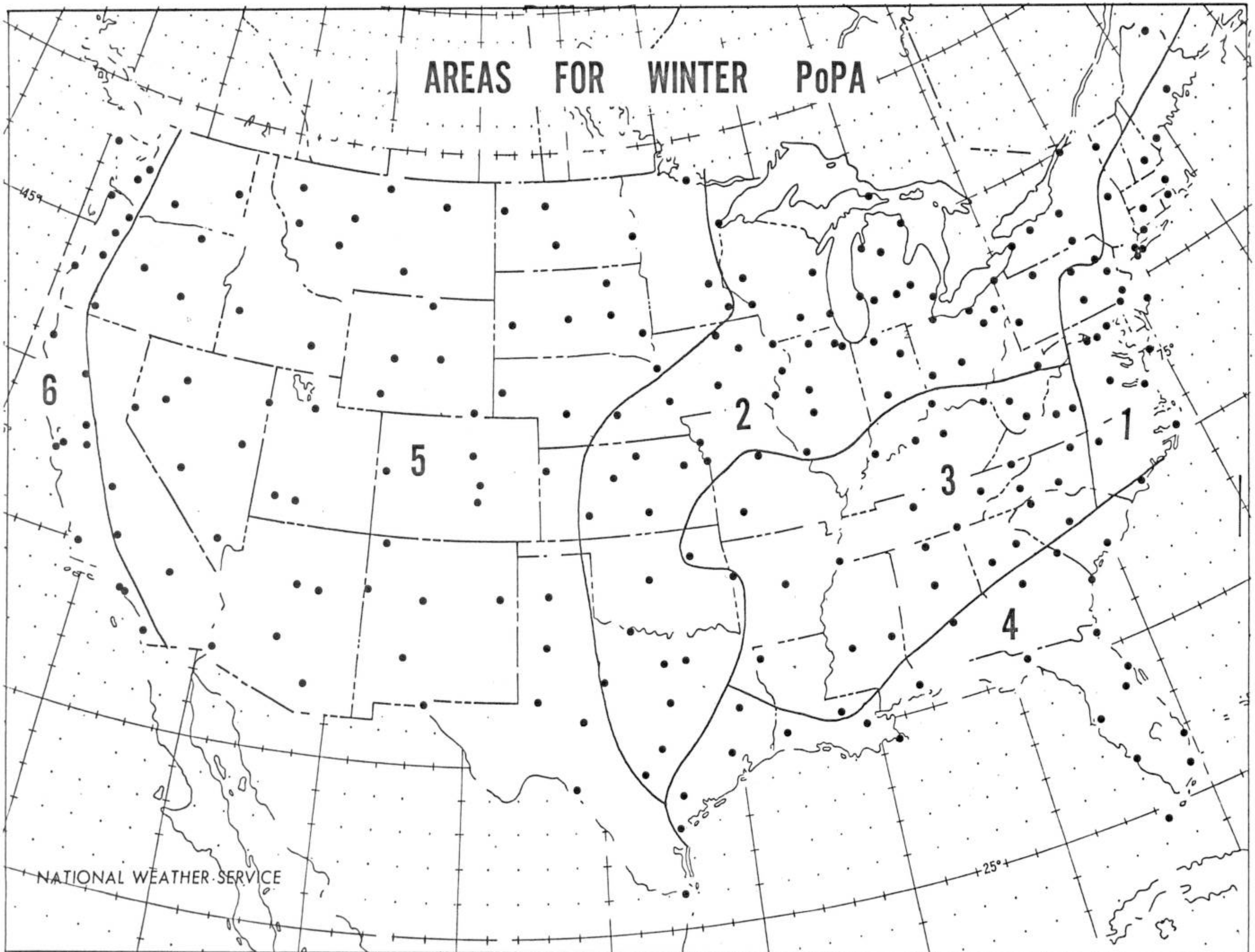

Figure 2. The 6 regions used for the POPA2 system during the winter season of 1974-1975.

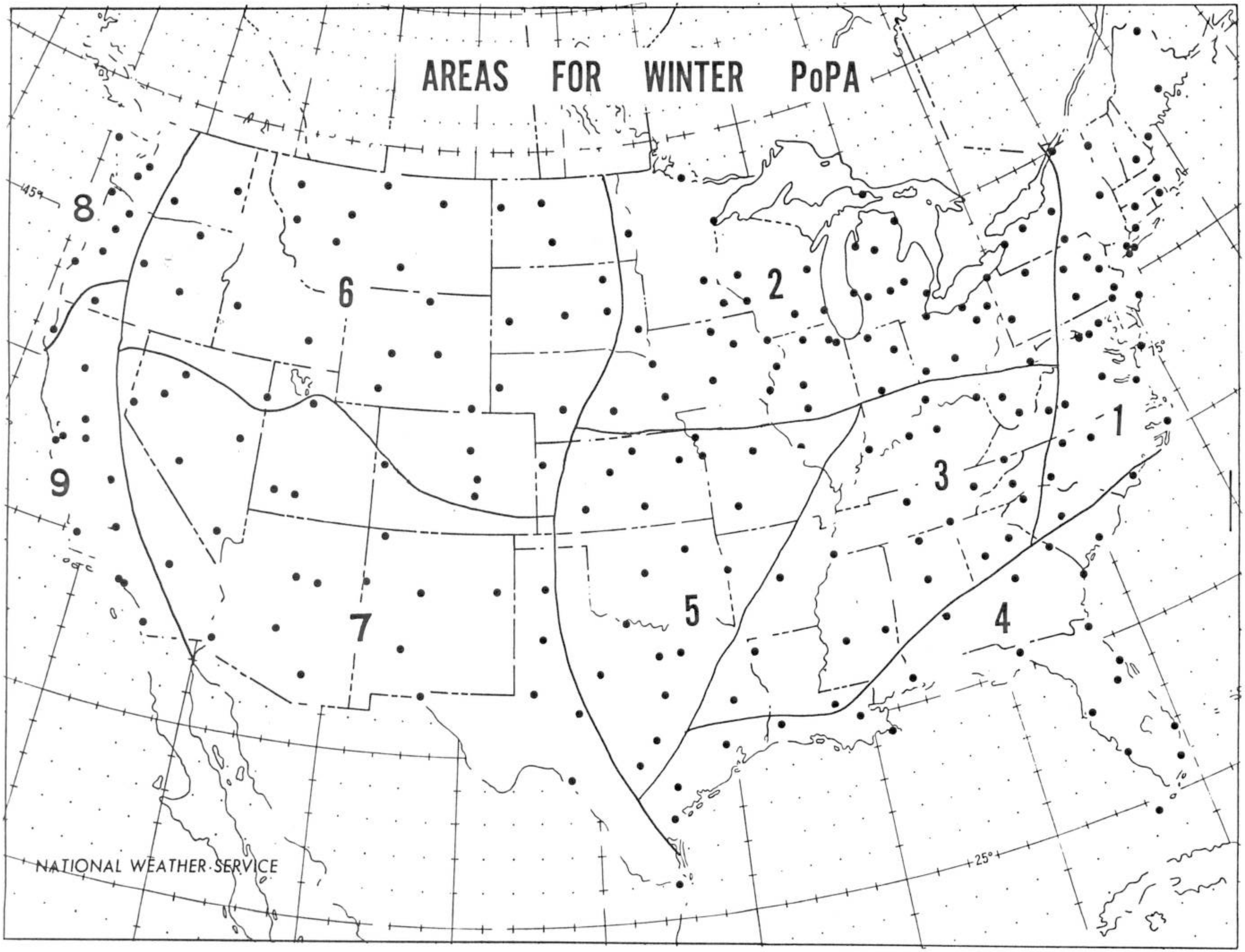

Figure 3. The 9 regions used for the POPA2 system during the winter season of 1975-1976.

Table 1. Verification of NMC subjective and 1974-1975 POPA2 system forecasts for the period February 19 - March 31, 1975, for the periods 12-36 hr after 0000 GMT for 39 days (A), 36-60 hr after 0000 GMT for 38 days (B), and 24-48 hr after 1200 GMT for 34 days (C).

(A)

Score	Category (in)							
	≥ .25		≥ .50		≥ 1.0		≥ 2.0	
	SUBJ	POPA	SUBJ	POPA	SUBJ	POPA	SUBJ	POPA
Threat Score	.399	.402	.313	.301	.215	.174	.138	.067
Bias	1.61	1.20	1.76	1.57	2.13	1.81	1.91	1.35

(B)

Score	Category (in)							
	≥ .25		≥ .50		≥ 1.0		≥ 2.0	
	SUBJ	POPA	SUBJ	POPA	SUBJ	POPA	SUBJ	POPA
Threat Score	.343	.300	.257	.186	.158	.100	.015	.010
Bias	1.48	1.02	1.57	1.03	1.61	1.29	.97	1.77

(C)

Score	Category (in)							
	≥ .25		≥ .50		≥ 1.0		≥ 2.0	
	SUBJ	POPA	SUBJ	POPA	SUBJ	POPA	SUBJ	POPA
Threat Score	.348	.349	.281	.258	.209	.165	.103	.058
Bias	1.61	1.53	1.66	2.01	1.84	2.67	1.74	3.66

Table 2. Verification of NMC subjective, 1974-1975 POPA2 system, and LFM forecasts for the period February 19 - March 31, 1975 for the category $\geq$ .25 inch for the periods 12-18 hr after 0000 GMT for 37 days (A), 18-24 hr after 0000 GMT for 37 days (B), 12-18 hr after 1200 GMT for 34 days (C), and 18-24 hr after 1200 GMT for 32 days (D).

(A)

Score	SUBJ	POPA	LFM
Threat Score	.259	.251	.257
Bias	1.84	1.57	1.92

(B)

Score	SUBJ	POPA	LFM
Threat Score	.158	.176	.212
Bias	2.25	1.97	2.47

(C)

Score	SUBJ	POPA	LFM
Threat Score	.184	.184	.223
Bias	1.78	1.96	1.80

(D)

Score	SUBJ	POPA	LFM
Threat Score	.168	.204	.175
Bias	1.52	1.94	2.00

Table 3. Verification of NMC subjective, 1974-1975 POPA2 system, and 1975-1976 POPA2 system forecasts for the periods 12-36 hr after 0000 GMT for 28 days (A) and 36-60 hr after 0000 GMT for 27 days (B) in March 1975.

(A)

Score	Category (in)											
	≥ .25			≥ .50			≥ 1.0			≥ 2.0		
	SUBJ	POPA 74-75	POPA 75-76	SUBJ	POPA 74-75	POPA 75-76	SUBJ	POPA 74-75	POPA 75-76	SUBJ	POPA 74-75	POPA 75-76
Threat Score	.389	.388	.411	.302	.296	.292	.235	.184	.187	.179	.077	.127
Bias	1.63	1.22	1.38	1.79	1.63	1.94	2.10	1.68	2.18	1.55	1.26	1.58

(B)

Score	Category (in)											
	≥ .25			≥ .50			≥ 1.0			≥ 2.0		
	SUBJ	POPA 74-75	POPA 75-76	SUBJ	POPA 74-75	POPA 75-76	SUBJ	POPA 74-75	POPA 75-76	SUBJ	POPA 74-75	POPA 75-76
Threat Score	.366	.305	.333	.275	.183	.197	.185	.083	.122	.016	.014	.017
Bias	1.52	1.07	1.42	1.69	1.06	1.58	1.74	1.28	1.03	1.00	1.35	0.97

Table 4. Verification of NMC subjective, 1974-1975 POPA2 system, 1975-1976 POPA2 system, and LFM forecasts for the category ≥ .25 inch for the period 12-18 hr after 0000 GMT for 26 days in March 1975.

Score	SUBJ	POPA 74-75	POPA 75-76	LFM
Threat Score	.231	.254	.242	.240
Bias	1.82	1.75	2.15	2.05